STATISTICS (STAT)

STAT 101 Introductory Business Statistics
Data summaries and descriptive statistics; introduction to a statistical computer package; Probability: distributions, expectation, variance, covariance, portfolios, central limit theorem; statistical inference of univariate data; Statistical inference for bivariate data: inference for intrinsically linear simple regression models. This course will have a business focus, but is not inappropriate for students in the college. This course may be taken concurrently with the prerequisite with the prerequisite with instructor permission.
One-term course offered either term
Prerequisite: MATH 104 OR MATH 110
Activity: Lecture
1.0 Course Unit

STAT 102 Introductory Business Statistics
Continuation of STAT 101. A thorough treatment of multiple regression, model selection, analysis of variance, linear logistic regression; introduction to time series. Business applications. This course may be taken concurrently with the prerequisite with the prerequisite with instructor permission.
One-term course offered either term
Prerequisite: STAT 101
Activity: Lecture
1.0 Course Unit

STAT 111 Introductory Statistics
Introduction to concepts in probability. Basic statistical inference procedures of estimation, confidence intervals and hypothesis testing directed towards applications in science and medicine. The use of the JMP statistical package. Knowledge of high school algebra is required for this course.
One-term course offered either term
Activity: Recitation
1.0 Course Unit

STAT 112 Introductory Statistics
Further development of the material in STAT 111, in particular the analysis of variance, multiple regression, non-parametric procedures and the analysis of categorical data. Data analysis via statistical packages. This course may be taken concurrently with the prerequisite with the prerequisite with instructor permission.
One-term course offered either term
Prerequisite: STAT 111
Activity: Lecture
1.0 Course Unit

STAT 399 Independent Study
Written permission of instructor and the department course coordinator required to enroll in this course.
One-term course offered either term
Activity: Independent Study
1.0 Course Unit

STAT 401 Sports Analytics: A Capstone Course
This course would introduce undergraduate students to the growing field of sports analytics, while allowing them to implement and integrate their knowledge base by exploring real sports data sets to solve real problems. While the context will be sports related, the skills and techniques gained will be widely applicable and generalizable with applications in diverse areas. Prerequisites: Must be a declared Statistics Concentrator or Business Analytics Concentrator or Statistics Minor or Data Science Minor. Permission from the Instructor is required.
Taught by: Abraham J. Wyner
Activity: Lecture
0.5 Course Units

STAT 405 Statistical Computing with R
The goal of this course is to introduce students to the R programming language and related eco-system. This course will provide a skill-set that is in demand in both the research and business environments. In addition, R is a platform that is used and required in other advanced classes taught at Wharton, so that this class will prepare students for these higher level classes and electives.
Taught by: Su, Waterman
One-term course offered either term
Also Offered As: STAT 705
Prerequisite: STAT 102 OR STAT 112 OR STAT 430
Activity: Lecture
0.5 Course Units

STAT 422 Predictive Analytics for Business
This course follows from the introductory regression classes, STAT 102, STAT 112, and STAT 431 for undergraduates and STAT 613 for MBAs. It extends the ideas from regression modeling, focusing on the core business task of predictive analytics as applied to realistic business related data sets. In particular it introduces automated model selection tools, such as stepwise regression and various current model selection criteria such as AIC and BIC. It delves into classification methodologies such as logistic regression. It also introduces classification and regression trees (CART) and the popular predictive methodology known as the random forest. By the end of the course the student will be familiar with and have applied all these tools and will be ready to use them in a work setting. The methodologies can all be implemented in either the JMP or R software packages. This course may be taken concurrently with the prerequisite with the prerequisite with instructor permission.
One-term course offered either term
Also Offered As: STAT 722
Prerequisite: STAT 102 OR STAT 112 OR STAT 431
Activity: Lecture
0.5 Course Units
STAT 424 Text Analytics
This course introduces methods for the analysis of unstructured data, focusing on statistical models for text. Techniques include those for sentiment analysis, topic models, and predictive analytics. Course includes topics from natural language processing (NLP), such as identifying parts of speech, parsing sentences (e.g., subject and predicate), and named entity recognition (people and places). Unsupervised techniques suited to feature creation provide variables suited to traditional statistical models (regression) and more recent approaches (regression trees). Examples that span the course illustrate the success of text analytics. Hierarchical generating models often associated with nonparametric Bayesian analysis supply theoretical foundations. Students should be familiar with regression models at the level of STAT 102 and the R statistics language at the level of STAT 405. Familiarity with the R-Studio development environment is presumed, as well as common R packages such as stringr, dplyr and ggplot. Those with more knowledge of Statistics, such as from STAT 422, or computing skills will benefit. The predominant software used in the course is R, with bits of JMP when helpful for interactive illustration. Familiarity with basic probability models is helpful but not presumed.

Taught by: Stine
One-term course offered either term
Also Offered As: STAT 724
Activity: Lecture
0.5 Course Units

STAT 430 Probability
Discrete and continuous sample spaces and probability; random variables, distributions, independence; expectation and generating functions; Markov chains and recurrence theory.

One-term course offered either term
Also Offered As: STAT 510
Prerequisite: MATH 114 OR MATH 115
Activity: Lecture
1.0 Course Unit

STAT 431 Statistical Inference
Graphical displays; one- and two-sample confidence intervals; one- and two-sample hypothesis tests; one- and two-way ANOVA; simple and multiple linear least-squares regression; nonlinear regression; variable selection; logistic regression; categorical data analysis; goodness-of-fit tests. A methodology course. This course does not have business applications but has significant overlap with STAT 101 and 102. This course may be taken concurrently with the prerequisite with instructor permission.

One-term course offered either term
Also Offered As: STAT 511
Prerequisite: STAT 430
Activity: Lecture
1.0 Course Unit

STAT 432 Mathematical Statistics
An introduction to the mathematical theory of statistics. Estimation, with a focus on properties of sufficient statistics and maximum likelihood estimators. Hypothesis testing, with a focus on likelihood ratio tests and the consequent development of "t" tests and hypothesis tests in regression and ANOVA. Nonparametric procedures. This course may be taken concurrently with the prerequisite with instructor permission.

Course usually offered in spring term
Also Offered As: STAT 512
Prerequisite: STAT 430 OR STAT 510
Activity: Lecture
1.0 Course Unit

STAT 433 Stochastic Processes
An introduction to Stochastic Processes. The primary focus is on Markov Chains, Martingales and Gaussian Processes. We will discuss many interesting applications from physics to economics. Topics may include: simulations of path functions, game theory and linear programming, stochastic optimization, Brownian Motion and Black-Scholes. This course may be taken concurrently with the prerequisite with instructor permission.

One-term course offered either term
Also Offered As: STAT 533
Prerequisite: STAT 430
Activity: Lecture
1.0 Course Unit

STAT 435 Forecasting Methods for Management
This course provides an introduction to the wide range of techniques available for statistical modelling and forecasting of time series. Regression methods for decomposition models, trends and seasonality, spectral analysis, distributed lag models, autoregressive-moving average modeling, forecasting, exponential smoothing, and ARCH and GARCH models will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations. This course may be taken concurrently with the prerequisite with instructor permission.

Taught by: Shaman
One-term course offered either term
Also Offered As: STAT 535, STAT 711
Prerequisite: STAT 102 OR STAT 112 OR STAT 431
Activity: Lecture
1.0 Course Unit

STAT 442 Introduction to Bayesian Data Analysis
The course will introduce data analysis from the Bayesian perspective to undergraduate students. We will cover important concepts in Bayesian probability modeling as well as estimation using both optimization and simulation-based strategies. Key topics covered in the course include hierarchical models, mixture models, hidden Markov models and Markov Chain Monte Carlo. A course in probability (STAT 430 or equivalent); a course in statistical inference (STAT 102, STAT 112, STAT 431 or equivalent); and experience with the statistical software R (at the level of STAT 405 or STAT 470) are recommended.

Taught by: Jensen
Course usually offered in spring term
Activity: Lecture
1.0 Course Unit
STAT 451 Fundamentals of Actuarial Science I
This course is the usual entry point in the actuarial science program. It is required for students who plan to concentrate or minor in actuarial science. It can also be taken by others interested in the mathematics of personal finance and the use of mortality tables. For future actuaries, it provides the necessary knowledge of compound interest and its applications, and basic life contingencies definition to be used throughout their studies. Non-actuaries will be introduced to practical applications of finance mathematics, such as loan amortization and bond pricing, and premium calculation of typical life insurance contracts. Main topics include annuities, loans and bonds; basic principles of life contingencies and determination of annuity and insurance benefits and premiums. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Lemaire
Course usually offered in fall term
Also Offered As: BEPP 451, BEPP 851, STAT 851
Prerequisite: MATH 104 AND STAT 430
Activity: Lecture
1.0 Course Unit

STAT 452 Fundamentals of Actuarial Science II
This specialized course is usually only taken by Wharton students who plan to concentrate in actuarial science and Penn students who plan to minor in actuarial mathematics. It provides a comprehensive analysis of advanced life contingencies problems such as reserving, multiple life functions, multiple decrement theory with application to the valuation of pension plans. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Lemaire
Course usually offered in spring term
Also Offered As: BEPP 452, BEPP 852, STAT 852
Prerequisite: STAT 451 OR BEPP 451
Activity: Lecture
1.0 Course Unit

STAT 453 Actuarial Statistics
This course covers models for insurer’s losses, and applications of Markov chains. Poisson processes, including extensions such as non-homogeneous, compound, and mixed Poisson processes are studied in detail. The compound model is then used to establish the distribution of losses. An extensive section on Markov chains provides the theory to forecast future states of the process, as well as numerous applications of Markov chains to insurance, finance, and genetics. The course is abundantly illustrated by examples from the insurance and finance literature. While most of the students taking the course are future actuaries, other students interested in applications of statistics may discover in class many fascinating applications of stochastic processes and Markov chains. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Lemaire
Course usually offered in fall term
Also Offered As: BEPP 453, BEPP 853, STAT 853
Prerequisite: STAT 430
Activity: Lecture
1.0 Course Unit

STAT 470 Data Analytics and Statistical Computing
This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests). Prerequisite: Waiving the Statistics Core completely if prerequisites are not met. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Johndrow
One-term course offered either term
Also Offered As: STAT 503, STAT 770
Prerequisite: (STAT 101 AND STAT 102) OR (STAT 111 AND STAT 112) OR STAT 431 OR (ECON 103 AND ECON 104)
Activity: Lecture
1.0 Course Unit

STAT 471 Modern Data Mining
Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Zhao
One-term course offered either term
Also Offered As: STAT 571, STAT 701
Prerequisite: STAT 102 OR STAT 112 OR STAT 431
Activity: Lecture
1.0 Course Unit

STAT 474 Modern Regression for the Social, Behavioral and Biological Sciences
Function estimation and data exploration using extensions of regression analysis: smoothers, semiparametric and nonparametric regression, and supervised machine learning. Conceptual foundations are addressed as well as hands-on use for data analysis. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Berk
Course usually offered in spring term
Also Offered As: CRIM 474, STAT 974
Prerequisite: STAT 102 OR STAT 112
Activity: Lecture
1.0 Course Unit
STAT 475 Sample Survey Design
This course will cover the design and analysis of sample surveys. Topics include simple sampling, stratified sampling, cluster sampling, graphics, regression analysis using complex surveys and methods for handling nonresponse bias. This course may be taken concurrently with the prerequisite with instructor permission.
Course not offered every year
Prerequisite: STAT 102 OR STAT 112 OR STAT 431
Activity: Lecture
1.0 Course Unit

STAT 476 Applied Probability Models in Marketing
This course will expose students to the theoretical and empirical "building blocks" that will allow them to construct, estimate, and interpret powerful models of consumer behavior. Over the years, researchers and practitioners have used these models for a wide variety of applications, such as new product sales, forecasting, analyses of media usage, and targeted marketing programs. Other disciplines have seen equally broad utilization of these techniques. The course will be entirely lecture-based with a strong emphasis on real-time problem solving. Most sessions will feature sophisticated numerical investigations using Microsoft Excel. Much of the material is highly technical.
Taught by: Fader
Course usually offered in spring term
Also Offered As: MKTG 476, MKTG 776, STAT 776
Activity: Lecture
1.0 Course Unit

STAT 477 Introduction to Python for Data Science
The goal of this course is to introduce the Python programming language within the context of the closely related areas of statistics and data science. Students will develop a solid grasp of Python programming basics, as they are exposed to the entire data science workflow, starting from interacting with SQL databases to query and retrieve data, through data wrangling, reshaping, summarizing, analyzing and ultimately reporting their results. Competency in Python is a critical skill for students interested in data science. Prerequisites: No prior programming experience is expected, but statistics, through the level of multiple regression is required. This requirement may be fulfilled with Undergraduate courses such as Stat 102, Stat 112.
Taught by: Richard Waterman
Also Offered As: OIDD 477, OIDD 777, STAT 777
Activity: Lecture
0.5 Course Units

STAT 480 Advanced Statistical Computing
This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation. Prerequisite: STAT 770 or 705 or equivalent background acquired through a combination of online courses that teach the R language and practical experience. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Buja
Course usually offered in spring term
Also Offered As: STAT 580, STAT 780
Prerequisite: STAT 405 OR STAT 470
Activity: Lecture
1.0 Course Unit

STAT 490 Causal Inference
Questions about cause are at the heart of many everyday decisions and public policies. Does eating an egg every day cause people to live longer or shorter or have no effect? Do gun control laws cause more or less murders or have no effect? Causal inference is the subfield of statistics that considers how we should make inferences about such questions. This course will cover the key concepts and methods of causal inference rigorously. The course is intended for statistics concentrators and minors. Knowledge of R such as that covered in STAT 405 or STAT 470 is recommended.
Taught by: Small
Course usually offered in spring term
Also Offered As: STAT 590
Prerequisite: STAT 430 AND (STAT 102 OR STAT 112 OR STAT 431)
Activity: Lecture
1.0 Course Unit

STAT 500 Applied Regression and Analysis of Variance
An applied graduate level course in multiple regression and analysis of variance for students who have completed an undergraduate course in basic statistical methods. Emphasis is on practical methods of data analysis and their interpretation. Covers model building, general linear hypothesis, residual analysis, leverage and influence, one-way anova, two-way anova, factorial anova. Primarily for doctoral students in the managerial, behavioral, social and health sciences. Permission of instructor required to enroll.
Taught by: Rosenbaum
Course usually offered in fall term
Also Offered As: BSTA 550, PSYC 611
Activity: Lecture
1.0 Course Unit

STAT 501 Introduction to Nonparametric Methods and Log-linear Models
An applied graduate level course for students who have completed an undergraduate course in basic statistical methods. Covers two unrelated topics: loglinear and logit models for discrete data and nonparametric methods for nonnormal data. Emphasis is on practical methods of data analysis and their interpretation. Primarily for doctoral students in the managerial, behavioral, social and health sciences. Permission of instructor required to enroll.
Taught by: Rosenbaum
Course usually offered in spring term
Also Offered As: PSYC 612
Activity: Lecture
1.0 Course Unit

STAT 503 Data Analytics and Statistical Computing
This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrapping, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests). Prerequisite: Two courses at the statistics 400 or 500 level.
Taught by: Buja
One-term course offered either term
Also Offered As: STAT 470, STAT 770
Activity: Lecture
1.0 Course Unit
STAT 510 Probability
One-term course offered either term
Also Offered As: STAT 430
Activity: Lecture
1.0 Course Unit

STAT 511 Statistical Inference
Graphical displays; one- and two-sample confidence intervals; one- and two-sample hypothesis tests; one- and two-way ANOVA; simple and multiple linear least-squares regression; nonlinear regression; variable selection; logistic regression; categorical data analysis; goodness-of-fit tests. A methodology course.
One-term course offered either term
Also Offered As: STAT 431
Prerequisite: STAT 510
Activity: Lecture
1.0 Course Unit

STAT 512 Mathematical Statistics
An introduction to the mathematical theory of statistics. Estimation, with a focus on properties of sufficient statistics and maximum likelihood estimators. Hypothesis testing, with a focus on likelihood ratio tests and the consequent development of "t" tests and hypothesis tests in regression and ANOVA. Nonparametric procedures.
Course usually offered in spring term
Also Offered As: STAT 432
Prerequisite: STAT 430 OR STAT 510
Activity: Lecture
1.0 Course Unit

STAT 515 Advanced Statistical Inference I
STAT 515 is aimed at first-year Ph.D. students and builds a good foundation in statistical inference from the first principles of probability.
Taught by: Krieger
Course usually offered in fall term
Prerequisite: STAT 510
Activity: Lecture
1.0 Course Unit

STAT 516 Advanced Statistical Inference II
STAT 516 is a natural continuation of STAT 515, and the main focus is on asymptotic evaluations and regression models. Time permitting, it also discusses some basic nonparametric statistical methods.
Taught by: Low
Course usually offered in spring term
Prerequisite: STAT 515
Activity: Lecture
1.0 Course Unit

STAT 520 Applied Econometrics I
This is a course in econometrics for graduate students. The goal is to prepare students for empirical research by studying econometric methodology and its theoretical foundations. Students taking the course should be familiar with elementary statistical methodology and basic linear algebra, and should have some programming experience. Topics include conditional expectation and linear projection, asymptotic statistical theory, ordinary least squares estimation, the bootstrap and jackknife, instrumental variables and two-stage least squares, specification tests, systems of equations, generalized least squares, and introduction to use of linear panel data models.
Taught by: Shaman
Course usually offered in fall term
Prerequisite: MATH 114 AND MATH 312
Activity: Lecture
1.0 Course Unit

STAT 521 Applied Econometrics II
Topics include system estimation with instrumental variables, fixed effects and random effects estimation, M-estimation, nonlinear regression, quantile regression, maximum likelihood estimation, generalized method of moments estimation, minimum distance estimation, and binary and multinomial response models. Both theory and applications will be stressed.
Taught by: Shaman
Course usually offered in spring term
Prerequisite: STAT 520
Activity: Lecture
1.0 Course Unit

STAT 533 Stochastic Processes
An introduction to Stochastic Processes. The primary focus is on Markov Chains, Martingales and Gaussian Processes. We will discuss many interesting applications from physics to economics. Topics may include: simulations of path functions, game theory and linear programming, stochastic optimization, Brownian Motion and Black-Scholes.
One-term course offered either term
Also Offered As: STAT 433
Prerequisite: STAT 510
Activity: Lecture
1.0 Course Unit

STAT 535 Forecasting Methods for Management
This course provides an introduction to the wide range of techniques available for statistical modelling and forecasting of time series. Regression methods for decomposition models, trends and seasonality, spectral analysis, distributed lag models, autoregressive-moving average modeling, forecasting, exponential smoothing, and ARCH and GARCH models will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Shaman
One-term course offered either term
Also Offered As: STAT 435, STAT 711
Prerequisite: (STAT 613 OR STAT 621) OR STAT 102
Activity: Lecture
1.0 Course Unit
STAT 542 Bayesian Methods and Computation
Sophisticated tools for probability modeling and data analysis from the Bayesian perspective. Hierarchical models, mixture models and Monte Carlo simulation techniques.
Taught by: Jensen
Course usually offered in spring term
Prerequisite: STAT 430 OR STAT 510
Activity: Lecture
1.0 Course Unit

STAT 571 Modern Data Mining
Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class. Prerequisite: two courses at the statistics 400 or 500 level or permission from instructor.
Taught by: Zhao
One-term course offered either term
Also Offered As: STAT 471, STAT 701
Activity: Lecture
1.0 Course Unit

STAT 580 Advanced Statistical Computing
This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation. Prerequisite: STAT 503, 705, or 770 or equivalent background acquired through a combination of online courses that teach the R language and practical experience.
Taught by: Buja
Course usually offered in spring term
Also Offered As: STAT 480, STAT 780
Prerequisite: STAT 503 OR STAT 705 OR STAT 770
Activity: Lecture
1.0 Course Unit

STAT 590 Causal Inference
Questions about cause are at the heart of many everyday decisions and public policies. Does eating an egg every day cause people to live longer or shorter or have no effect? Do gun control laws cause more or less murders or have no effect? Causal inference is the subfield of statistics that considers how we should make inferences about such questions. This course will cover the key concepts and methods of causal inference rigorously. Background in probability and statistics; some knowledge of R is recommended.
Taught by: Small
Course usually offered in spring term
Also Offered As: STAT 490
Activity: Lecture
1.0 Course Unit

STAT 613 Regression Analysis for Business
This course provides the fundamental methods of statistical analysis, the art and science if extracting information from data. The course will begin with a focus on the basic elements of exploratory data analysis, probability theory and statistical inference. With this as a foundation, it will proceed to explore the use of the key statistical methodology known as regression analysis for solving business problems, such as the prediction of future sales and the response of the market to price changes. The use of regression diagnostics and various graphical displays supplement the basic numerical summaries and provides insight into the validity of the models. Specific important topics covered include least squares estimation, residuals and outliers, tests and confidence intervals, correlation and autocorrelation, collinearity, and randomization. The presentation relies upon computer software for most of the needed calculations, and the resulting style focuses on construction of models, interpretation of results, and critical evaluation of assumptions.
Course usually offered in fall term
Prerequisite: STAT 611
Activity: Lecture
1.0 Course Unit

STAT 621 Accelerated Regression Analysis for Business
STAT 621 is intended for students with recent, practical knowledge of the use of regression analysis in the context of business applications. This course covers the material of STAT 613, but omits the foundations to focus on regression modeling. The course reviews statistical hypothesis testing and confidence intervals for the sake of standardizing terminology and introducing software, and then moves into regression modeling. The pace presumes recent exposure to both the theory and practice of regression and will not be accommodating to students who have not seen or used these methods previously. The interpretation of regression models within the context of applications will be stressed, presuming knowledge of the underlying assumptions and derivations. The scope of regression modeling that is covered includes multiple regression analysis with categorical effects, regression diagnostic procedures, interactions, and time series structure. The presentation of the course relies on computer software that will be introduced in the initial lectures. Recent exposure to the theory and practice of regression modeling is recommended.
Taught by: George
Course usually offered in fall term
Activity: Lecture
0.5 Course Units
STAT 701 Modern Data Mining
Modern Data Mining: Statistics or Data Science has been evolving rapidly to keep up with the modern world. While classical multiple regression and logistic regression technique continue to be the major tools we go beyond to include methods built on top of linear models such as LASSO and Ridge regression. Contemporary methods such as KNN (K nearest neighbor), Random Forest, Support Vector Machines, Principal Component Analyses (PCA), the bootstrap and others are also covered. Text mining especially through PCA is another topic of the course. While learning all the techniques, we keep in mind that our goal is to tackle real problems. Not only do we go through a large collection of interesting, challenging real-life data sets but we also learn how to use the free, powerful software "R" in connection with each of the methods exposed in the class. Prerequisite: two courses at the statistics 400 or 500 level or permission from instructor.
Taught by: Zhao
One-term course offered either term
Also Offered As: STAT 471, STAT 571
Activity: Lecture
1.0 Course Unit

STAT 705 Statistical Computing with R
The goal of this course is to introduce students to the R programming language and related eco-system. This course will provide a skill-set that is in demand in both the research and business environments. In addition, R is a platform that is used and required in other advanced classes taught at Wharton, so that this class will prepare students for these higher level classes and electives.
Taught by: Su, Waterman
One-term course offered either term
Also Offered As: STAT 405
Prerequisite: STAT 613 OR STAT 621
Activity: Lecture
0.5 Course Units

STAT 711 Forecasting Methods for Management
This course provides an introduction to the wide range of techniques available for statistical modelling and forecasting of time series. Regression methods for decomposition models, trends and seasonality, spectral analysis, distributed lag models, autoregressive-moving average modeling, forecasting, exponential smoothing, and ARCH and GARCH models will be surveyed. The emphasis will be on applications, rather than technical foundations and derivations. The techniques will be studied critically, with examination of their usefulness and limitations. This course may be taken concurrently with the prerequisite with instructor permission.
Taught by: Shaman
One-term course offered either term
Also Offered As: STAT 435, STAT 535
Prerequisite: (STAT 613 OR STAT 621) OR STAT 102
Activity: Lecture
1.0 Course Unit

STAT 722 Predictive Analytics for Business
This course follows from the introductory regression classes, STAT 102, STAT 112, and STAT 431 for undergraduates and STAT 613 for MBAs. It extends the ideas from regression modeling, focusing on the core business task of predictive analytics as applied to realistic business related data sets. In particular it introduces automated model selection tools, such as stepwise regression and various current model selection criteria such as AIC and BIC. It delves into classification methodologies such as logistic regression. It also introduces classification and regression trees (CART) and the popular predictive methodology known as the random forest. By the end of the course the student will be familiar with and have applied all these tools and will be ready to use them in a work setting. The methodologies can all be implemented in either the JMP or R software packages. This course is formerly STAT 622.
One-term course offered either term
Also Offered As: STAT 422
Prerequisite: STAT 613 OR STAT 621
Activity: Lecture
0.5 Course Units

STAT 724 Text Analytics
This course introduces methods for the analysis of unstructured data, focusing on statistical models for text. Techniques include those for sentiment analysis, topic models, and predictive analytics. Course includes topics from natural language processing (NLP), such as identifying parts of speech, parsing sentences (e.g., subject and predicate), and named entity recognition (people and places). Unsupervised techniques suited to feature creation provide variables suited to traditional statistical models (regression) and more recent approaches (regression trees). Examples that span the course illustrate the success of text analytics. Hierarchical generating models often associated with nonparametric Bayesian analysis supply theoretical foundations. Students should be familiar with regression models at the level of STAT 613 and the R statistics language at the level of STAT 705. Familiarity with the R-Studio development environment is presumed, as well as common R packages such as stringr, dplyr and ggplot. Those with more knowledge of Statistics, such as from STAT 722, or computing skills will benefit. The predominant software used in the course is R, with bits of JMP when helpful for interactive illustration. Familiarity with basic probability models is helpful but not presumed.
Taught by: Stine
One-term course offered either term
Also Offered As: STAT 424
Activity: Lecture
0.5 Course Units

STAT 770 Data Analytics and Statistical Computing
This course will introduce a high-level programming language, called R, that is widely used for statistical data analysis. Using R, we will study and practice the following methodologies: data cleaning, feature extraction; web scrubbing, text analysis; data visualization; fitting statistical models; simulation of probability distributions and statistical models; statistical inference methods that use simulations (bootstrap, permutation tests). Prerequisite: Two courses at the statistics 400 or 500 level.
Taught by: Buja
One-term course offered either term
Also Offered As: STAT 470, STAT 503
Activity: Lecture
1.0 Course Unit
STAT 776 Applied Probability Models in Marketing
This course will expose students to the theoretical and empirical “building blocks” that will allow them to construct, estimate, and interpret powerful models of consumer behavior. Over the years, researchers and practitioners have used these models for a wide variety of applications, such as new product sales, forecasting, analyses of media usage, and targeted marketing programs. Other disciplines have seen equally broad utilization of these techniques. The course will be entirely lecture-based with a strong emphasis on real-time problem solving. Most sessions will feature sophisticated numerical investigations using Microsoft Excel. Much of the material is highly technical.
Taught by: Fader
Course usually offered in spring term
Also Offered As: MKTG 476, MKTG 776, STAT 476
Activity: Lecture
1.0 Course Unit
Notes: Format: Lecture, real-time problem solving

STAT 777 Introduction to Python for Data Science
The goal of this course is to introduce the Python programming language within the context of the closely related areas of statistics and data science. Students will develop a solid grasp of Python programming basics, as they are exposed to the entire data science workflow, starting from interacting with SQL databases to query and retrieve data, through data wrangling, reshaping, summarizing, analyzing and ultimately reporting their results. Competency in Python is a critical skill for students interested in data science. Prerequisites: No prior programming experience is expected, but statistics, through the level of multiple regression is required. This requirement may be fulfilled with MBA courses such as STAT 613/621; or by waiving MBA statistics.
Taught by: Richard Waterman
Also Offered As: OIDD 477, OIDD 777, STAT 477
Activity: Lecture
0.5 Course Units

STAT 780 Advanced Statistical Computing
This course will build on the fundamental concepts introduced in the prerequisite courses to allow students to acquire knowledge and programming skills in large-scale data analysis, data visualization, and stochastic simulation. Prerequisite: STAT 503, 705, or 770 or equivalent background acquired through a combination of online courses that teach the R language and practical experience.
Taught by: Buja
Course usually offered in spring term
Also Offered As: STAT 480, STAT 580
Prerequisite: STAT 503 OR STAT 705 OR STAT 770
Activity: Lecture
1.0 Course Unit

STAT 786 Applied Probability Models in Marketing
This course is the usual entry point in the actuarial science program. It is required for students who plan to concentrate or minor in actuarial science. It can also be taken by others interested in the mathematics of personal finance and the use of mortality tables. For future actuaries, it provides the necessary knowledge of compound interest and its applications, and basic life contingencies definition to be used throughout their studies. Non-actuaries will be introduced to practical applications of finance mathematics, such as loan amortization and bond pricing, and premium calculation of typical life insurance contracts. Main topics include annuities, loans and bonds; basic principles of life contingencies and determination of annuity and insurance benefits and premiums. Prerequisite: One semester of calculus.
Taught by: Lemaire
Course usually offered in fall term
Also Offered As: BEPP 451, BEPP 851, STAT 451
Activity: Lecture
1.0 Course Unit

STAT 852 Fundamentals of Actuarial Science II
This specialized course is usually only taken by Wharton students who plan to concentrate in actuarial science and Penn students who plan to minor in actuarial mathematics. It provides a comprehensive analysis of advanced life contingencies problems such as reserving, multiple life functions, multiple decrement theory with application to the valuation of pension plans.
Taught by: Lemaire
Course usually offered in spring term
Also Offered As: BEPP 452, BEPP 852, STAT 452
Prerequisite: STAT 851 OR BEPP 851
Activity: Lecture
1.0 Course Unit

STAT 853 Actuarial Statistics
This course covers models for insurer’s losses, and applications of Markov chains. Poisson processes, including extensions such as non-homogeneous, compound, and mixed Poisson processes are studied in detail. The compound model is then used to establish the distribution of losses. An extensive section on Markov chains provides the theory to forecast future states of the process, as well as numerous applications of Markov chains to insurance, finance, and genetics. The course is abundantly illustrated by examples from the insurance and finance literature. While most of the students taking the course are future actuaries, other students interested in applications of statistics may discover in class many fascinating applications of stochastic processes and Markov chains. Prerequisite: Two semesters of statistics.
Taught by: Lemaire
Course usually offered in fall term
Also Offered As: BEPP 453, BEPP 853, STAT 453
Activity: Lecture
1.0 Course Unit

STAT 899 Independent Study
Written permission of instructor, the department MBA advisor and course coordinator required to enroll.
One-term course offered either term
Activity: Independent Study
1.0 Course Unit
STAT 915 Nonparametric Inference
Statistical inference when the functional form of the distribution is not specified. Nonparametric function estimation, density estimation, survival analysis, contingency tables, association, and efficiency.
Course not offered every year
Prerequisite: STAT 520
Activity: Lecture
1.0 Course Unit

STAT 920 Sample Survey Methods
This course will cover the design and analysis of sample surveys. Topics include simple random sampling, stratified sampling, cluster sampling, graphics, regression analysis using complex surveys and methods for handling nonresponse bias.
Taught by: Small
Course not offered every year
Prerequisite: STAT 520 OR STAT 961 OR STAT 970
Activity: Lecture
1.0 Course Unit

STAT 921 Observational Studies
This course will cover statistical methods for the design and analysis of observational studies. Topics will include the potential outcomes framework for causal inference; randomized experiments; matching and propensity score methods for controlling confounding in observational studies; tests of hidden bias; sensitivity analysis; and instrumental variables.
Taught by: Small
One-term course offered either term
Prerequisite: STAT 520 OR STAT 961 OR STAT 970
Activity: Lecture
1.0 Course Unit

STAT 925 Multivariate Analysis: Theory
This is a course that prepares PhD students in statistics for research in multivariate statistics and high dimensional statistical inference. Topics from classical multivariate statistics include the multivariate normal distribution and the Wishart distribution; estimation and hypothesis testing of mean vectors and covariance matrices; principal component analysis, canonical correlation analysis and discriminant analysis; etc. Topics from modern multivariate statistics include the Marcenko-Pastur law, the Tracy-Widom law, nonparametric estimation and hypothesis testing of high-dimensional covariance matrices, high-dimensional principal component analysis, etc.
Taught by: Ma
Course not offered every year
Prerequisite: STAT 930 OR STAT 970 OR STAT 972
Activity: Lecture
1.0 Course Unit

STAT 926 Multivariate Analysis: Methodology
This is a course that prepares PhD students in statistics for research in multivariate statistics and data visualization. The emphasis will be on a deep conceptual understanding of multivariate methods to the point where students will propose variations and extensions to existing methods or whole new approaches to problems previously solved by classical methods. Topics include: principal component analysis, canonical correlation analysis, generalized canonical analysis; nonlinear extensions of multivariate methods based on optimal transformations of quantitative variables and optimal scaling of categorical variables; shrinkage- and sparsity-based extensions to classical methods; clustering methods of the k-means and hierarchical varieties; multidimensional scaling, graph drawing, and manifold estimation.
Taught by: Buja
Course not offered every year
Prerequisite: STAT 961
Activity: Lecture
1.0 Course Unit

STAT 927 Bayesian Statistical Theory and Methods
This graduate course will cover the modeling and computation required to perform advanced data analysis from the Bayesian perspective. We will cover fundamental topics in Bayesian probability modeling and implementation, including recent advances in both optimization and simulation-based estimation strategies. Key topics covered in the course include hierarchical and mixture models, Markov Chain Monte Carlo, hidden Markov and dynamic linear models, tree models, Gaussian processes and nonparametric Bayesian strategies.
Taught by: Jensen
Course not offered every year
Prerequisite: STAT 430 OR STAT 510
Activity: Lecture
1.0 Course Unit

STAT 928 Statistical Learning Theory
Statistical learning theory studies the statistical aspects of machine learning and automated reasoning, through the use of (sampled) data. In particular, the focus is on characterizing the generalization ability of learning algorithms in terms of how well they perform on "new" data when trained on some given data set. The focus of the course is on: providing the fundamental tools used in this analysis; understanding the performance of widely used learning algorithms; understanding the "art" of designing good algorithms, both in terms of statistical and computational properties. Potential topics include: empirical process theory; online learning; stochastic optimization; margin based algorithms; feature selection; concentration of measure. Background in probability and linear algebra recommended.
Course usually offered in fall term
Activity: Lecture
1.0 Course Unit

STAT 930 Probability Theory
Taught by: Pemantle
Course usually offered in fall term
Also Offered As: MATH 648
Prerequisite: STAT 430 OR STAT 510 OR MATH 608
Activity: Lecture
1.0 Course Unit
STAT 931 Stochastic Processes
Also Offered As: MATH 649
Prerequisite: MATH 648 OR STAT 930
Activity: Lecture
1.0 Course Unit

STAT 955 Stochastic Calculus and Financial Applications
Selected topics in the theory of probability and stochastic processes.
Course usually offered in fall term
Prerequisite: STAT 930
Activity: Lecture
1.0 Course Unit

STAT 960 Statistical Algorithms and Computation
This course aims to prepare students for graduate work in the design, analysis, and implementation of statistical algorithms. The target audience is Ph.D. students in statistics or in adjacent fields, such as computer science, mathematics, electrical engineering, computational biology, economics, and marketing. We will take a fundamental approach and focus on classes of algorithms of primary importance in statistics and statistical machine learning. Some meta-clases of algorithms that may receive significant attention are optimization, sampling, and numerical linear algebra. I aim to make the content complementary rather than overlapping with other courses at Penn, such as ESE605, CIS677, and the CIS700 series. While there may be some overlap in the portions of the course that cover optimization, the sampling (Monte Carlo and related) aspects of the course are, to my knowledge, hard to find elsewhere at Penn. The course is fast paced and I expect a certain degree of mathematical preparation. Most students in the above mentioned programs will have the requisite mathematics background. I also expect familiarity with an appropriate programming language such as R, python, or matlab. The course will be mostly language agnostic. However, I may at times give example code in one of these languages, and you will be expected to be able to read the code even if it is not in your "primary" language. We may make use of various open-source toolboxes and packages for these environments, such as the Stan probabilistic programming language (best used with R) and the cvx toolbox for convex programming (available for multiple platforms but perhaps best used with matlab).
Taught by: James Johndrow
Course usually offered in spring term
Prerequisites: STAT 930 and STAT 961
Activity: Lecture
1.0 Course Unit

STAT 961 Statistical Methodology
This is a course that prepares 1st year PhD students in statistics for a research career. This is not an applied statistics course. Topics covered include: linear models and their high-dimensional geometry, statistical inference illustrated with linear models, diagnostics for linear models, bootstrap and permutation inference, principal component analysis, smoothing and cross-validation.
Taught by: Buja
Course usually offered in fall term
Prerequisite: STAT 431 OR STAT 520
Activity: Lecture
1.0 Course Unit

STAT 962 Advanced Methods for Applied Statistics
This course is designed for Ph.D. students in statistics and will cover various advanced methods and models that are useful in applied statistics. Topics for the course will include missing data, measurement error, nonlinear and generalized linear regression models, survival analysis, experimental design, longitudinal studies, building R packages and reproducible research.
Taught by: Small
Course usually offered in spring term
Prerequisite: STAT 961
Activity: Lecture
1.0 Course Unit

STAT 970 Mathematical Statistics
Decision theory and statistical optimality criteria, sufficiency, point estimation and hypothesis testing methods and theory.
Taught by: Small
Course usually offered in fall term
Prerequisite: STAT 431 OR STAT 520
Activity: Lecture
1.0 Course Unit

STAT 971 Introduction to Linear Statistical Models
Taught by: Ma
Course usually offered in spring term
Prerequisite: STAT 970
Activity: Lecture
1.0 Course Unit

STAT 972 Advanced Topics in Mathematical Statistics
A continuation of STAT 970.
Taught by: Cai
One-term course offered either term
Prerequisite: STAT 970 AND STAT 971
Activity: Lecture
1.0 Course Unit

STAT 974 Modern Regression for the Social, Behavioral and Biological Sciences
Function estimation and data exploration using extensions of regression analysis: smoothers, semiparametric and nonparametric regression, and supervised machine learning. Conceptual foundations are addressed as well as hands-on use for data analysis.
Taught by: Berk
Course usually offered in spring term
Also Offered As: CRIM 474, STAT 474
Prerequisite: STAT 102 OR STAT 112
Activity: Lecture
1.0 Course Unit

STAT 991 Seminar in Advanced Application of Statistics
This seminar will be taken by doctoral candidates after the completion of most of their coursework. Topics vary from year to year and are chosen from advance probability, statistical inference, robust methods, and decision theory with principal emphasis on applications.
One-term course offered either term
Activity: Seminar
1.0 Course Unit
STAT 995 Dissertation
One-term course offered either term
Activity: Dissertation
1.0 Course Unit

STAT 999 Independent Study
Written permission of instructor and the department course coordinator required to enroll.
One-term course offered either term
Activity: Independent Study
1.0 Course Unit